Fermi National Accelerator Laboratory

Top Quark Kinematics and Mass Determination

H.H. Williams
For the CDF Collaboration

Fermi National Accelerator Laboratory
P.O. Box 500, Batavia, Illinois 60510

University of Pennsylvania

October 1994

Disclaimer

Top Quark Kinematics and Mass Determination

H.H. Williams

University of Pennsylvania

On behalf of the CDF Collaboration

Abstract

An analysis is presented of 10 W + $\geq$ 3 jet events, each with evidence for the presence of a b quark, that were recently observed by the CDF collaboration. Seven of these events include a fourth jet and can be explicitly reconstructed as $t\bar{t}$ production. The best estimate of the top quark mass is $M_t = 174 \pm 10^{+13}_{-12}$ GeV/c^2. A study has also been performed to see if the *kinematical* properties of events with W + $\geq$ 3 jets gives evidence for top production. An excess of events with large jet energies, compared to that expected from direct production of W + $\geq$ 3 jets, is observed. A large fraction of these events also contain a b-quark and a fourth jet.

1. Introduction

In a previous paper [1] and talk presented at this conference [2], evidence for $t\bar{t}$ production from the Collider Detector at Fermilab (CDF) experiment was presented based on the observation of events with two high P_T leptons and missing $E_t(\not{E}_t)$ and W + $\geq$ 3 jet events with a b quark. The presence of the b quark was indicated by the existence of either a secondary vertex or an additional, low P_T lepton. In this paper we present a determination of the top quark mass using the W + jet + b events which include a fourth jet, and also present additional evidence for top production based on kinematics of observed W + jet events.

2. Mass Determination

Selecting events with E_t (electron) > 20 GeV (or P_T(muon) > 20 GeV), $\not{E}_t$ > 20 GeV, and three jets with E_t > 15 GeV, $|\eta|$ < 2.0 yields a sample of 52 W + $\geq$ 3 jet events; 10 of these events exhibit a b quark tag, with three of the events having two tags. † The probability that these events, plus the two observed dilepton events, are produced by a background fluctuation is < 0.26%. To investigate whether these events are consistent with being produced by $t\bar{t}$ production and to determine to what extent the value of M_t may be determined, we require the presence of a fourth jet so that a one-to-one correspondence may be made between each jet and one of the partons in the reaction

$$t\bar{t} \rightarrow l\nu\bar{b} + q\bar{q}b \qquad (1)$$

To obtain a higher acceptance, we require the fourth jet to satisfy E_t > 8 GeV, $|\eta|$ < 2.4; Monte Carlo studies indicate that for $M_t = 170$ GeV 86% of $t\bar{t}$ events which have three jets with E_t > 15 GeV, $|\eta|$ < 2.0 will also have a fourth jet with E_t > 8 GeV, $|\eta|$ < 2.4 while only 60% will have a fourth jet with E_t > 15 GeV, $|\eta|$ < 2.0.

† For the *selection* of this sample of events, which is described in detail in [1], the jet energies and missing transverse energy, $\not{E}_t$, have not been corrected for η and ϕ dependent variations in the calorimetry response or for the energy which is outside the cone size $R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} = 0.4$.

Consistent with this expectation 7 of the 10 observed events have a fourth jet. The estimated background in this sample of events is $1.4^{+2.0}_{-1.1}$.

We associate each of the observed jets with one of the final state quarks in the reaction:

$$p\bar{p} \rightarrow t\bar{t}X \rightarrow Wb\,W\bar{b}\,X$$
$$\rightarrow l\nu b\,q\bar{q}\bar{b}\,X \qquad (2)$$

If a jet has a b tag, then it is required to be one of the b quarks. There are then six different possible assignments of the jets to the quarks, and in addition, there are two possible solutions for the longitudinal momentum of the neutrino. The measured energies of each jet are corrected for detector effects and an additional correction is made to estimate the energy of the original parton; a different correction factor is used depending on whether the jet is assigned to a light quark from the W or to a b quark, and on whether the jet has a soft lepton tag or not. The E_t of X is taken to be the total observed transverse energy not associated with the jets or lepton (multiplied by a correction factor of 1.6), but the longitudinal momentum and effective mass of X are left as free parameters. We also require $M_t = M_{\bar{t}}$ and that the decay products of the W's reconstruct to M_W. The resulting fit is overconstrained (2C) so that M_t may be determined. Of the twelve different possible fits for each event, the one with the lowest χ^2 is chosen.

We have tested this procedure by applying it to $t\bar{t}$ events generated with the Herwig program [3] and passed through a simulation of the CDF detector followed by the same reconstruction used for data. The distribution of reconstructed masses for events generated with $M_t = 170$ GeV is shown in Figure 1; the central value is reconstructed at 168 GeV/c^2 and the σ is 23 GeV/c^2.

For comparison, the distribution obtained if one always uses the correct parton-jet assignments is shown by the dashed line. We find that the fitting procedure leads to the correct parton-jet assignments in only 31% of events; however, while for events with incorrect assignments the reconstructed distribution is significantly wider than that shown by the dashed curve, the central value is not significantly altered. In comparison, Figure 2 shows the mass distribution of W $+ \geq 3$ jet events, generated according to the leading order matrix elements [4] and reconstructed as if they were $t\bar{t}$; the distribution is peaked at significantly lower values, approximately 140 GeV.

Monte Carlo studies indicate that 94% of top events should yield a good fit with $\chi^2 < 10$. However, W $+$ jet events also yield a good fit 83% of the time because of the large number of combinations of jet assignments; thus it is not possible to discriminate effectively between $t\bar{t}$ and W $+$ jets on the basis of there existing a good fit.
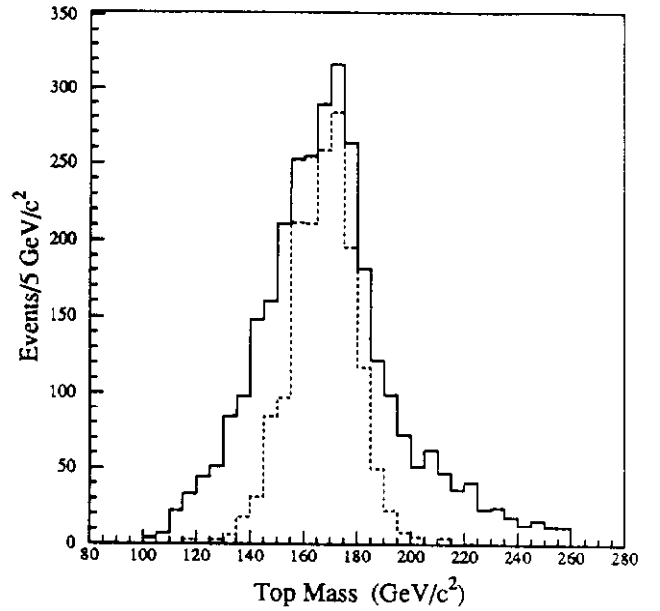

Figure 1. Reconstructed top mass distribution for Monte Carlo events generated with $M_{top} = 170$ GeV/c^2. The full histogram corresponds to the best fit obtained by the fitting program when requiring that one of the b jets is a b in the fit. The dashed histogram refers to the fit with the correct assignment for each of the jets.

The masses reconstructed for the 7 data events are shown by the solid line in Figure 3; the distribution bears more resemblance to that expected for top with M_t approximately 170 GeV than it does to the distribution for W $+$ jets.

To determine the best value of the top mass from the seven events, and to check whether the observed distribution is consistent with that expected, we have performed a fit to the seven events using a superposition of the mass distribution, $f_s(m, M_t)$, expected for top (Figure 1) and that expected for the W $+$ jet background, $f_b(m)$, (Figure 2). The likelihood function, L, is defined as

$$L = \frac{1}{\sqrt{2\pi}\sigma_b} e^{-\frac{(n_b - N_b)^2}{2\sigma_b^2}} \cdot \frac{e^{-(n_s + n_b)} \cdot (n_s + n_b)^N}{N!}$$
$$\prod_{i=1}^{N} \frac{(n_b f_b(m_i) + n_s f_s(m_i, M_t))}{(n_b + n_s)} \qquad (3)$$

where n_b and n_s represent the number of background and signal events, N is the number of observed events (7), N_b is the estimated number of background events (1.4), and σ_b is the background uncertainty, set equal to 1.6 for a Gaussian approximation to the asymmetric errors quoted above. This likelihood function is evaluated for several different values of M_t; the results are shown in the insert in Figure 3.

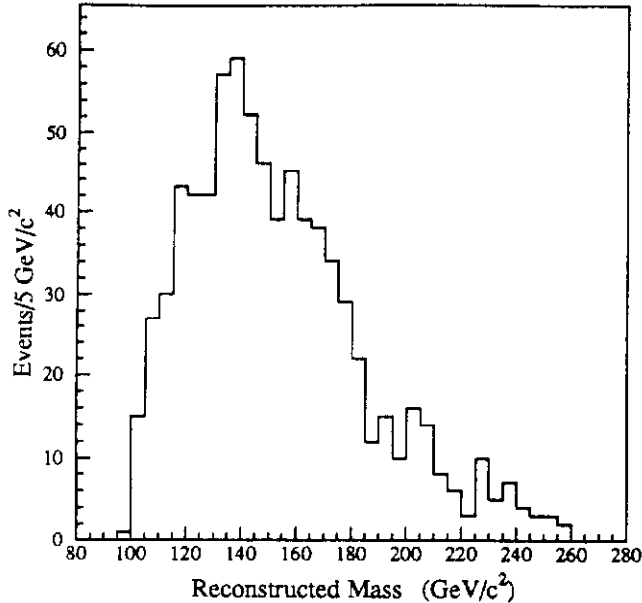The minimum value of L occurs for $M_t = 174$

Figure 2. Reconstructed mass distribution for W + multijet Monte Carlo events interpreted as $t\bar{t}$.
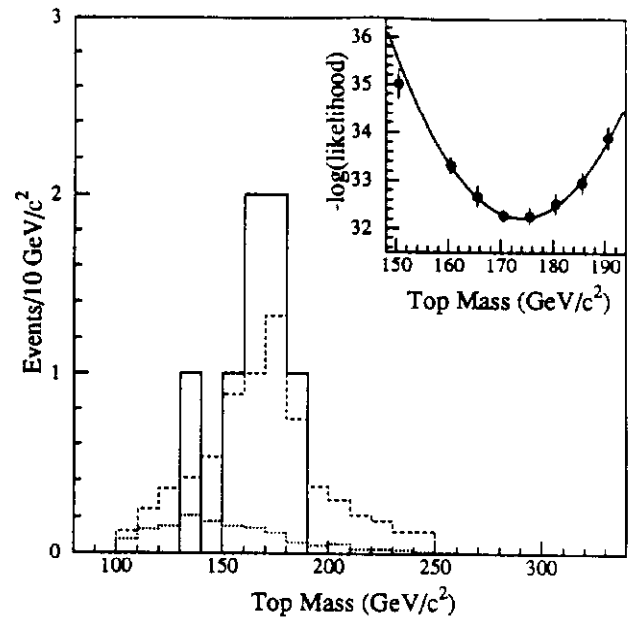

Figure 3. Top mass distribution for the data (solid histogram), the W + jets background (dots), and the sum of background + Monte Carlo $t\bar{t}$ for $M_{top} = 175$ GeV/c^2 (dashed). The background distribution is normalized to the 1.4 background events expected in the mass-fit sample. The inset shows the likelihood of the fit as a function of top mass.

GeV, and the statistical error, based on an increase in $\ln L$ of 0.5 and taking into account statistical errors in the Monte Carlo, is 10 GeV. We have checked whether the value of σ_{M_t} obtained is reasonable by performing a large number of seven event Monte Carlo experiments, each with the number of background events, n_b, generated according to a Poisson distribution with mean 1.4 ± 1.6. The masses, m, of the background events are distributed according to $f_b(m)$ and the 7 - n_b signal events are distributed according to $f_s(m, M_t)$. The most probable error is about 8–12 GeV in good agreement with the value of 10 GeV observed for this experiment. The likelihood observed for the fit to the data is also in good agreement with that expected from the Monte Carlo experiments.

We have studied systematic effects in the determination of M_t including (1) uncertainty in the energy scale of the calorimeter, estimated to range from 10% at 8 GeV to 3% at 100 GeV, (2) uncertainty in the correction for energy outside the cone of the jet, taken to be ± 10 %, (3) uncertainties in the shape of the background, (4) biases in the jet energies due to the tagging algorithms, and (5) variations due to different fitting procedures. The results are summarized in Table 1, and the uncertainties are added in quadrature to yield a total systematic error on M_t of $^{+13}_{-12}$ GeV/c^2.

	Systematic uncertainties	(%)
a.	Jet Energy Scale (detector effects)	1.8
b.	Gluon radiation effects on parton energy	4.4
c.	Different backgrounds	$^{+5.3}_{-4.4}$
d.	Effects due to tagging algorithms	1.4
e.	Different likelihood fits	1.1

Table 1. Systematic uncertainties in the top mass measurement.

Thus our best determination of the top quark mass from these events is:

$$M_t = 174 \pm 10^{+13}_{-12} \text{GeV}/c^2 \qquad (4)$$

3. Event Kinematics

We have also investigated whether the *kinematic* properties of the events, without requiring a b tag, yield evidence for $t\bar{t}$ production [5]. Alternatively, one may ask what cuts might provide a substantially enriched sample of top events. For this study we correct the jet energies for detector non-uniformities and energy outside the cone of the jets (before final selection of the events) to enable the most accurate comparison with the predicted kinematic distributions for W + jets. We require $E_t^c > 25$ GeV, where E_t^c is the missing E_t after the jet energies are corrected, and also add a requirement on the transverse mass of the W, $M_T > 40$ GeV/c^2 to
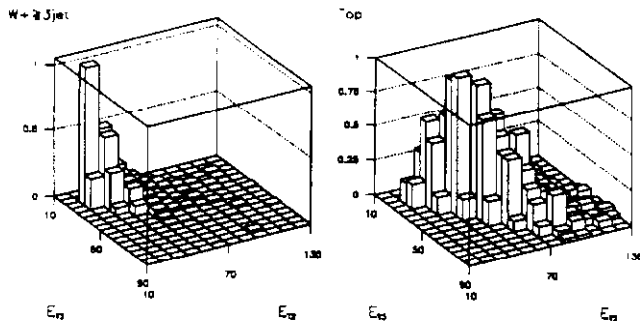
Figure 4. $d\sigma^2/dE_{T2}dE_{T3}$ for QCD W + $\geq$ 3 jet and top ($M_{top} = 170$ GeV/c^2) Monte Carlo events.

Figure 5. E_T distributions of W+$\geq$2 jet data (points) and Vecbos W + $\geq$ 2 jet events (histograms). (a) leading jet, (b) second to leading jet, (c,d) E_T distributions for events with $|\cos\theta^*_{max}| < 0.7$.

further minimize the background from non-W events. The jets are required to have $E_t > 20$ GeV and $|\eta| < 2.0$, and we also require a minimum separation between the jets, $|\delta R| > 0.7$, to minimize uncertainties in infra-red divergences in the calculated rates.

We have investigated a number of kinematic variables, including E_{t1}, E_{t2}, E_{t3} (the energies of the jets with the highest, second highest, and third highest energies), $\cos\theta^*$ (the angle of each jet with respect to the beam direction in the center of mass system of the jets + lepton + W), the aplanarity, and the existence and properties of a fourth jet. We find that E_{t2} and E_{t3} are among the most powerful variables for improving the signal-to-noise ratio in a W + $\geq$ 3 jet sample of events; this is indicated in Figure 4 which shows the expected distribution of events, as a function of these variables, for both W + jet production and for the production of $t\bar{t}$. The former process has been simulated using the VECBOS program [4], described in more detail below, while the latter process is simulated using HERWIG. Requirement of a fourth jet with transverse energy greater than 10–15 GeV is also a powerful discriminant between W + jet events and $t\bar{t}$ production; however, to maintain the maximum signal size and to minimize systematic errors, we initially do not make this requirement.

Another variable which is useful to improve the signal to background is θ^*, the angle of the jets with respect to the beam direction in the center of momentum frame of the jets and W; as the longitudinal momentum of the neutrino is unknown, we assume it to be zero in computing $\cos\theta^*$. The $\cos\theta^*$ distribution is expected to be significantly more peaked in the forward direction for W + jets than for $t\bar{t}$. We define a "signal" or "top- enriched" sample by requiring $|\cos\theta^*_{max}| \leq 0.7$ and a "control" or "background enriched" sample with $|\cos\theta^*_{max}| > 0.7$; the variable $|\cos\theta^*_{max}|$ is the largest value of $|\cos\theta^*|$ for any of the three jets. The samples contain 15 and 31 events respectively. On an *a priori* basis these two samples are expected to contain
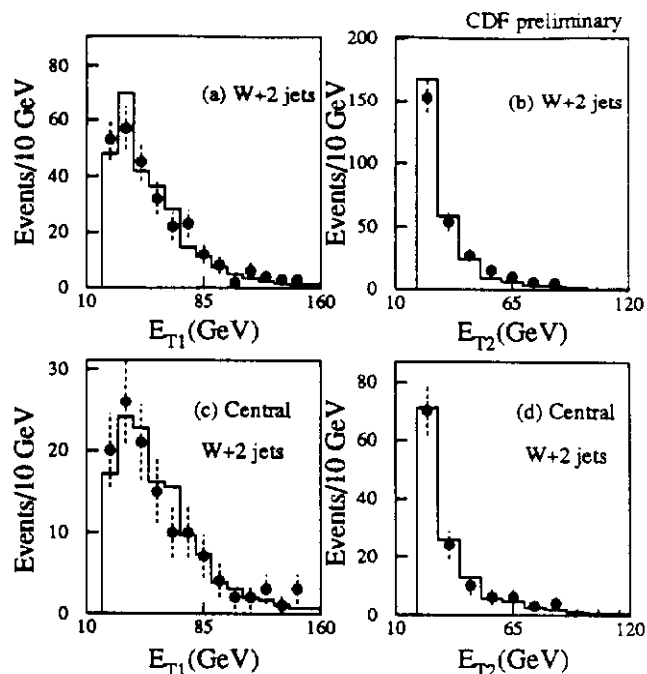
approximately equal numbers of top events, but the background enriched sample is expected to contain two–three times as many W + jet events.

3.1. W + jet Events: Predictions and Observations

A calculation of the properties of W + jet events produced via "standard" QCD interactions has existed for several years at the level of the leading order matrix elements at tree level for production of a W with n final state partons [4]. These matrix elements have been incorporated into the program Vecbos to allow generation of W events with $n = 0, 1, 2, 3, 4$ partons. In order to avoid infrared divergences, cuts are applied in the event generation requiring $P_T(\text{parton}) > 10$ GeV, $|\eta(\text{parton})| < 3.5$, and $|\delta R(\text{parton} - \text{parton})| > 0.4$. Two different fragmentation models, one of which includes simple fragmentation *a la* Field and Feynman [6] and one which includes parton evolution and fragmentation *a la* Herwig, have been utilized; they give very similar results. Numerous previous tests have indicated good agreement between the predictions and experimental observations [7]. Perhaps the best test is provided by the recent CDF $W+ \geq 2$ jet data sample. A comparison of the predicted and observed E_{t1} and E_{t2} distributions, presented in Figure 5, indicates quite good agreement. For this comparison, the Q^2 scale for
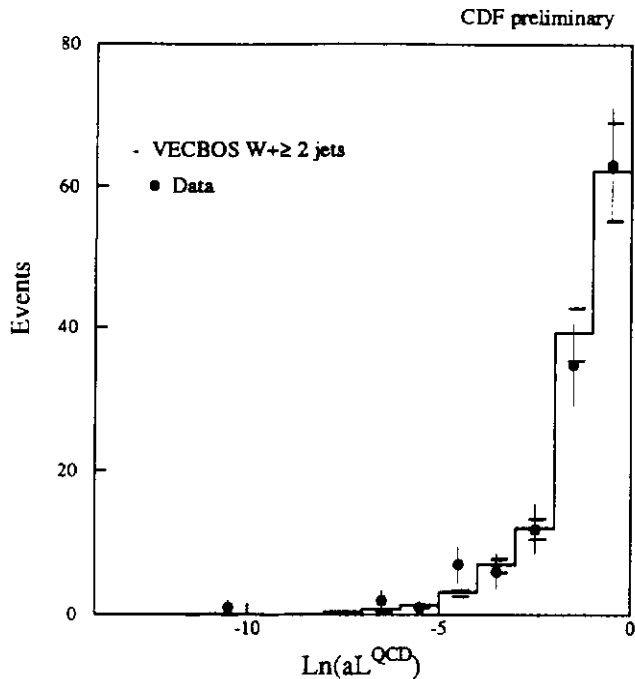
Figure 6. Comparison of the predicted (solid line) and observed (data points) $\ln(aL^{QCD})$ distributions (see text) for $W+ \geq 2$ jet events.

Figure 7. Jet energy distributions for the 15 events passing the signal sample selection cuts. There is one overflow in $E_T(\mathrm{jet}_1)$.

α_s is chosen to be $< P_t >^2$, the square of the average value of the P_t of the outgoing partons; this scale is also used for comparison of the predicted and observed distributions for three jet events. Use of the scale M_W^2 which yields slightly "harder" distributions, gives similar results. A comparison of the predicted and observed energy distributions for jets in $Z+2$ and $Z+3$ jet events also shows good agreement.

Before discussing the $W+3$ jet sample we introduce a variable which conveniently allows representation of the likelihood that an event with a given E_{t1} and E_{t2} is consistent with the expected parent distribution. The *absolute likelihood* is defined as

$$aL = 1/\sigma(d\sigma/dE_{t1}) \times 1/\sigma(d\sigma/dE_{t2}) \qquad (5)$$

We utilize a factorized product of the E_{t1} and E_{t2} distributions for simplicity and have verified that the absence of correlations does not significantly affect the analysis. The predicted and observed distributions of aL are shown in Figure 6. Events with relatively low jet energies, near the peak of the energy distributions, have fairly large likelihoods ($\ln(aL) > -3$ to -2) while events with large energies, on the tail of the energy distributions, have small likelihoods ($\ln(aL) \leq -5$ to -4). Again reasonably good agreement is obtained.

We now compare the observed (Figure 7) and predicted (Figure 8, dashed line) energy distributions for the events in the "signal enriched" sample of $W + \geq 3$
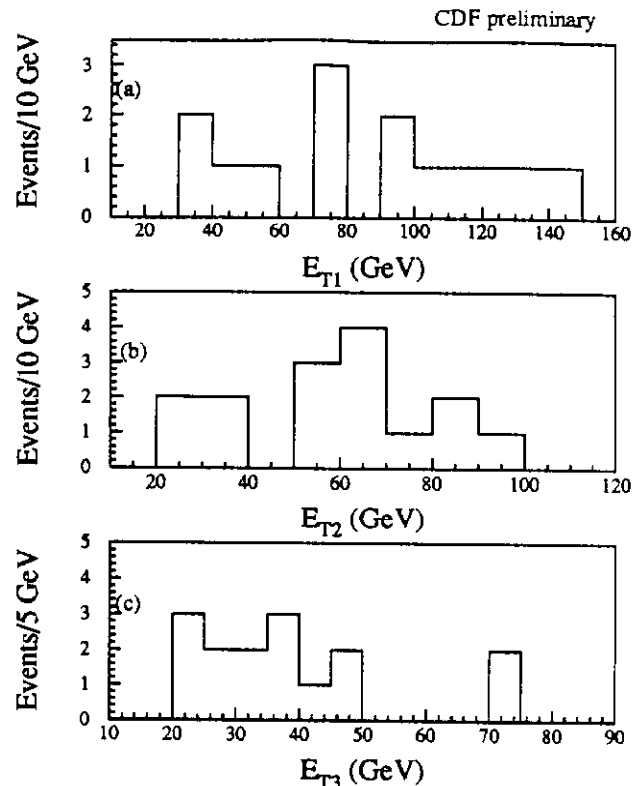
jet events; it appears that the data has a significantly larger number of events at high energies than would be expected. The data is consistent with a significant fraction of the events coming from $t\bar{t}$ production with $M_t \approx 170$ GeV; the E_t distributions for this process are also shown in Figure 8 (solid line).

To test quantitatively the consistency of the data with the $W + \geq 3$ jet expectations, combining the information from both E_{t2} and E_{t3}, we define an absolute likelihood for three jet events in an analogous fashion to that for two jet events:

$$aL^{QCD} = 1/\sigma(d\sigma/dE_{t2})1/\sigma(d\sigma/dE_{t3}) \qquad (6)$$

The expected and observed $\ln(aL^{QCD})$ distributions for the $W + \geq 3$ jet enriched sample (control sample) are shown in Figure 9; the data again agrees reasonably well with the expected distribution. However, the observed distribution for the signal or top-enriched sample, where the ratio of top to $W +$ jet events is expected on *a priori* grounds to be between 1:2 and 1:1, does not agree well with the expectations as is shown in Figure 10.

There is a clear excess of events at small likelihood ($aL^{QCD} < -6$). One may ask what the distribution of
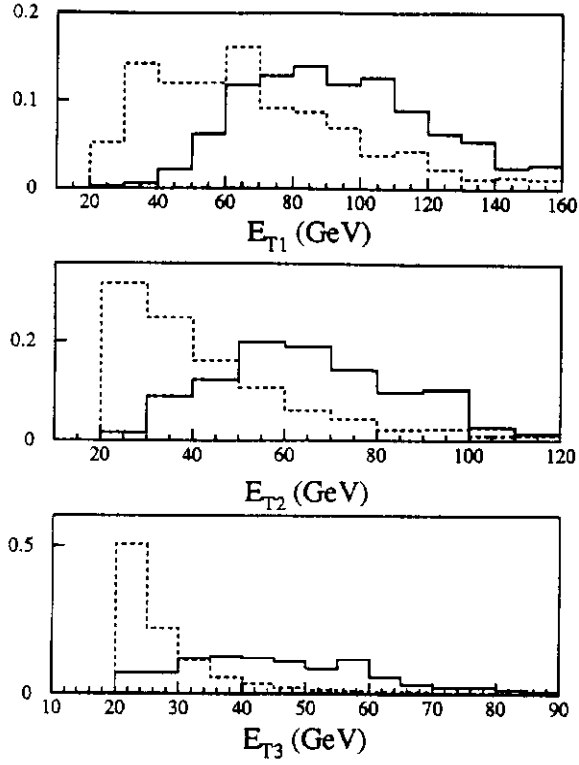
Figure 8. Jet energy distributions for Isajet $t\bar{t}$ (solid line) and Vecbos W + $\geq$ 3 jet events (dashed line) passing the signal sample selection cuts.
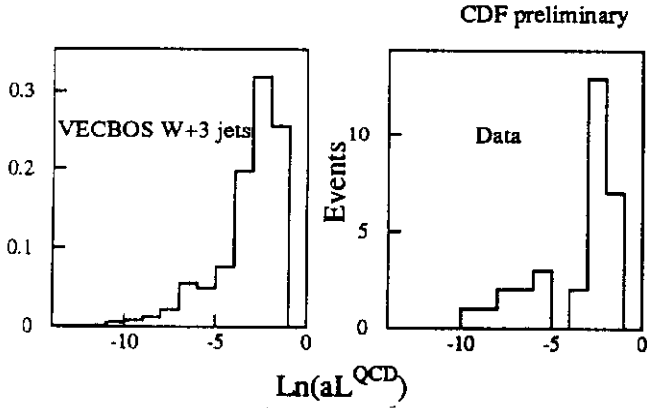

Figure 9. Distributions of $\ln(aL^{QCD})$ for W + $\geq$ 3 jet events with $|\cos(\theta^*_{max})| > 0.7$ (control sample); Vecbos Monte Carlo (left) and data (right).

top events would be in terms of the variable aL^{QCD}; this is shown in Figure 11 for a range of top masses.

A convenient way to represent the *relative likelihood* that an event is from W + $\geq$ 3 jets or top production is
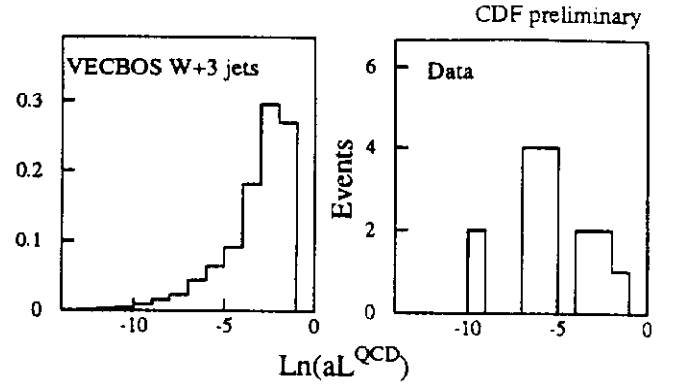

Figure 10. Distributions of $\ln(aL^{QCD})$ for W + $\geq$ 3 jet events with $|\cos(\theta^*_{max})| < 0.7$ (signal sample); Vecbos Monte Carlo (left) and data (right).
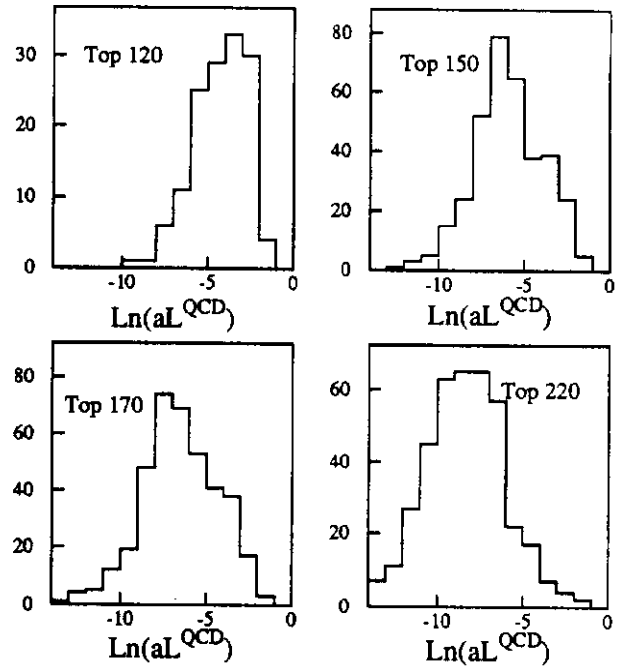

Figure 11. Expected distributions of Isajet $t\bar{t}$ events as a function of $\ln(aL)^{QCD}$, for a number of assumed top masses.

to define the quantity:

$$rL = aL^{Top}/aL^{QCD} \qquad (7)$$

The variable rL combines information from E_{t2} and E_{t3} and is equivalent to drawing contours in the E_{t2}-E_{t3} plane to select events predominantly from one process
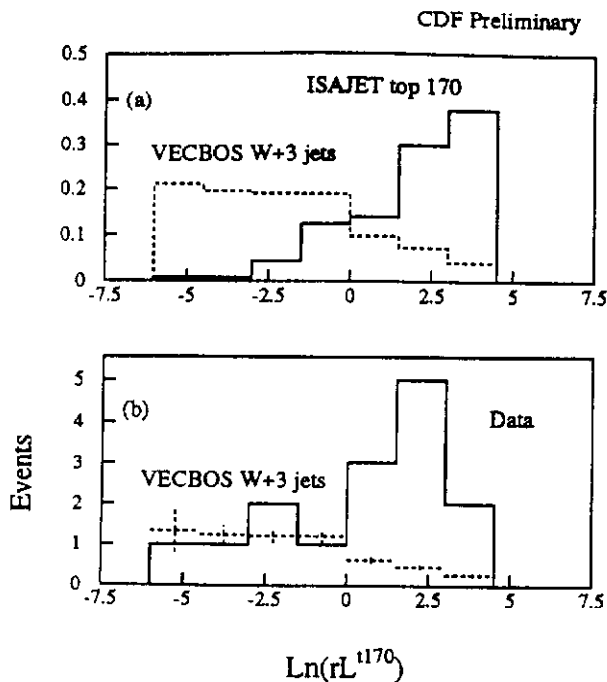
Figure 12. $\text{Ln}(rL^{t170})$ for QCD Vecbos, top Isajet ($M_{top} = 170$ GeV/c^2) and data events for W+ $\geq$ 3 central jet events (signal enriched sample) (a): W+$\geq$3 jet Vecbos (dotted histogram) and top Isajet (solid histogram), normalized to 1. (b): data (solid histogram) and Vecbos (dotted histogram). Vecbos is normalized to data in the region $\ln(rL) < 0$.

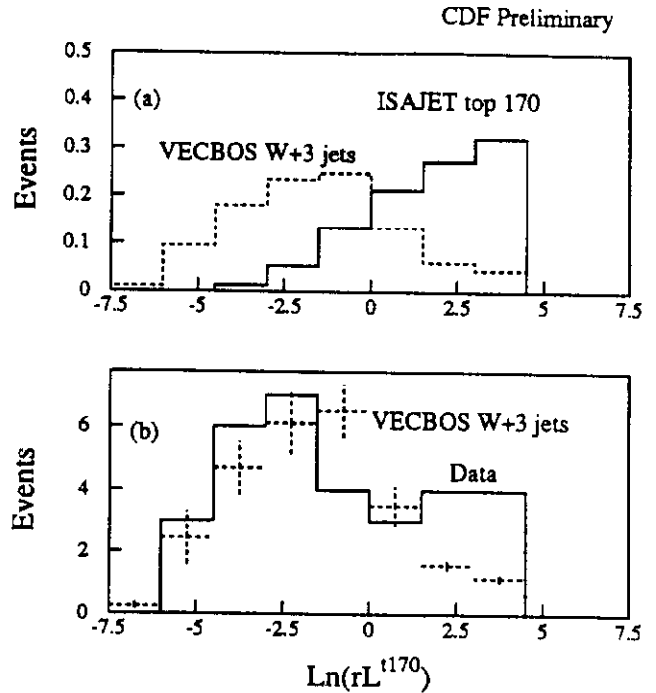Figure 13. Distributions of $\ln(rL)$ for the control sample. (a) Distributions of Vecbos W+3 jet (dotted histogram) and Isajet top events (solid histogram), normalized to 1. (b) 31 data events (solid histogram) versus Vecbos (dotted, with statistical errors). Vecbos has been normalized to data in the region $\ln(rL) < 0$.

or the other (as motivated by Figure 4); in a sample with equal numbers of W and $t\bar{t}$, events with $\ln(rL) > 0$ are most likely to be from $t\bar{t}$ production while those with $\ln(rL) < 0$ are most likely to be from W + $\geq$ 3 jet production. In defining this relative likelihood, it is of course necessary to choose a top mass in order to determine the expected E_{t2}, E_{t3} distributions. As indicated in Figure 11 the results are not very sensitive to the particular mass chosen; we choose $M_t = 170$ GeV in accordance with the results reported in the first half of this paper and the results from studies of electroweak interactions.

Figure 12 (a) shows the expected $\ln(rL)$ distributions for W + $\geq$ 3 jet events, and for $t\bar{t}$ production with $M_t = 170$ GeV as generated with the ISAJET program. Figure 12(b) shows the observed distribution; there is a clear excess of events at $\ln(rL) > 0$ over what would be expected from W + $\geq$ 3 jets alone.

One may compute the probability that the observed data is a fluctuation of the distribution expected for

W + $\geq$ 3 jet events from the binomial probability that a sample of 15 events, distributed according to the dashed curve in Figure 12 (a), yields 10 or more events with $\ln(rL) > 0$; the probability, before consideration of systematic effects, is small. We have tested the sensitivity of the result to various systematic effects—changing the energy scale of the calorimeter by $\pm$ 10% when calculating the expected distributions, assuming different values for the Q^2 scale for α_s in the Vecbos calculation, and using different assumptions for the fragmentation of the outgoing partons. These tests do not yield a significant change in the fraction of QCD produced W + jet events expected with $\ln(rL) > 0$. Quantitative evaluation of the probability that the observed distribution is a background fluctuation, including systematic effects and variation in the cuts, will be reported in the near future.

In the control sample, shown in Figure 13, the number of events with $\ln(rL) > 0$ is consistent with the presence of $t\bar{t}$ production, but any excess over that expected from W + $\geq$ 3 jets is not statistically significant.

In the next section, we study the presence of b tags to determine whether or not the excess of events at large $\ln(rL)$ in the signal sample are most likely due to top production.
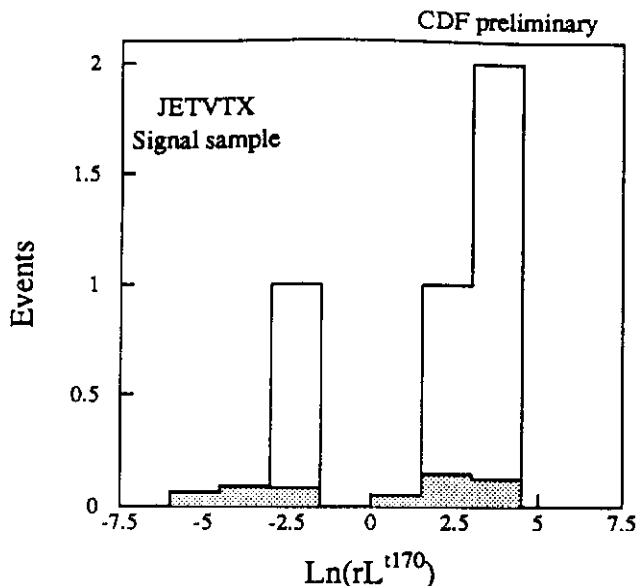
Figure 14. Distribution in ln(rL) of the 4 events of the signal sample tagged by the JETVTX vertexing algorithm. The expected fakes are shown as the shaded histogram.

3.2. b-tags

As mentioned above and described in [1, 2], evidence for b-quarks in the final state can be provided by the existence of a reconstructed secondary vertex or by the presence of a "soft" lepton in the event. Four of the 15 events in the signal sample have an SVX tag and their distribution as a function of $\ln(rL)$ is shown in Figure 14. The estimated number of tags expected in the absence of top production is shown by the shaded region. Similarly 4 of the 15 events include a soft-lepton tag (SLT) indicating the presence of a b-tag. The expected number of SLT tags in the absence of top production is of order one event.

3.3. Presence of a 4th jet

If the events with $\ln(rL) > 0$) are due to $t\bar{t}$, one would also expect a substantial fraction to include a fourth jet: according to the Herwig Monte Carlo approximately 80% of top events will include a fourth jet with $E_t > 15$ GeV (corrected energy) while the expected fraction for W $+ \geq 3$ jet events is much lower, approximately 30% due to the small value of α_s. We find that 7 of the 10 W $+ \geq 3$ jet events with $\ln(rL) > 0$ have a 4th jet while none of the 5 events with $\ln rL < 0$ do. Of these seven events, three have an SVX b-tag and three have a soft lepton tag. A W + 4 jet Monte Carlo, normalized so that it predicts 5 W $+ \geq 3$ jet events with $\ln rL < 0$, predicts of order 1 W + 4 jet event with $\ln rL > 0$. The fact that seven of the 10 W $+ \geq 3$ jet events with $\ln(rL) > 0$ have a fourth jet, and that there

is a total of six b tags among these seven events, is a strong confirmation that these events are indeed from $t\bar{t}$ production.

4. Conclusion

In a search for the top quark, [1, 2], the CDF experiment has observed 10 W $+ \geq 3$ jet events in which a secondary vertex or low P_T lepton indicates the presence of a b-quark. Three events exhibit two such b-quark tags. The experiment has also observed two events with two high P_T leptons and large missing E_T. The probability that this observation is due to a background fluctuation is < 0.26%. Seven of the ten W$+ \geq 3$ jet events include a fourth jet, and a detailed fit indicates that each of these events is consistent with coming from $t\bar{t}$ production. The best estimate of the top quark mass, based on these events, is $M_t = 174 \pm 10^{+13}_{-12}$ GeV/c^2.

We have also searched for evidence of top quark production based on the kinematic properties of W$+ \geq 3$ jet events without requiring a b-tag. For this study we impose slightly different selection criteria using corrected jet energies. We divide the sample into a signal-enriched (top quark) and background enriched (W$+ \geq 3$ jet) sample and define a relative likelihood (rL) that determines whether a given event fits better the $t\bar{t}$ or W $+ \geq 3$ jet hypothesis. We find that the background-enriched sample and a W $+ \geq 2$ jet sample are consistent with the expectations for W $+$ n jet events, while the signal sample has an excess of events with large jet energies (large relative likelihood for top production). Five of the ten events at large rL contain a b quark tag. Seven of the events also include a fourth jet. The presence of the b quark and the fourth jet provide strong supporting evidence that the excess of events at large rL are indeed from $t\bar{t}$ production. The 5-10 fold increase in data from the present run should allow a more precise determination of the top quark mass and a refinement of the techniques for selecting enriched samples of top quark events.

References

[1] F. Abe et al., Phys. Rev. Lett. 73 (1994) 225; F. Abe et al., Phys. Rev. D50, 2966 (1994).

[2] F. Bedeschi, " Evidence for Top Quark Production from the CDF Experiment", this proceedings.

[3] G. Marchesini and B.R. Webber, Nucl. Phys. B310 (1988) 461; G. Marchesini et al., Comput. Phys. Comm. 67, (1992) 465.

[4] F.A. Berends, W.T. Giele, H. Kuif, B. Tausk, Fermilab-Pub-90/213-T. W. Giele, Ph.D thesis, unpublished.

[5] A more detailed discussion can be found in M. Cobal, PhD thesis, Univ Pisa, INFN PI/AE/94/004 and M. Cobal, H. Grassman, S. Leone, Il Nuovo Cimento 107A, 75 (1994).

[6] R. Field and R. Feynman, Nucl. Phys. B136, (1978) 1

[7] F. Abe et. al, Phys. Rev. Lett. 70, 4042 (1993). J. Alitti et al, Phys. Lett. 268B, 145 (1992).